PENNSTATE

METROLOGY FOR HIGH ASSURANCE INFORMATION BROKERING

NIST WORKSHOP: ADVANCING MEASUREMENTS AND TESTING FOR INFORMATION TECHNOLOGY

Oct 26-27, 1998

Dr. Shashi Phoha
Head, Information Science & Technology Division
Assistant Director of Applied Research Laboratory, and
Professor of Electrical & Computer Engineering
sxp26@psu.edu
(814)863-8005

ELECTRONIC MEMORY

A multi-user mechanism which is capable of resolving global queries with differing and overlapping information needs, spatial scalability and temporal assumptions by drawing upon heterogeneous sources of dispersed and transient information embedded, possibly, in autonomous legacy systems.

EXAMPLES

- Web-based search engines
- Domain specific information brokers
- Enterprise-wide information brokers
- Collaboration Tools; Blackboard applications

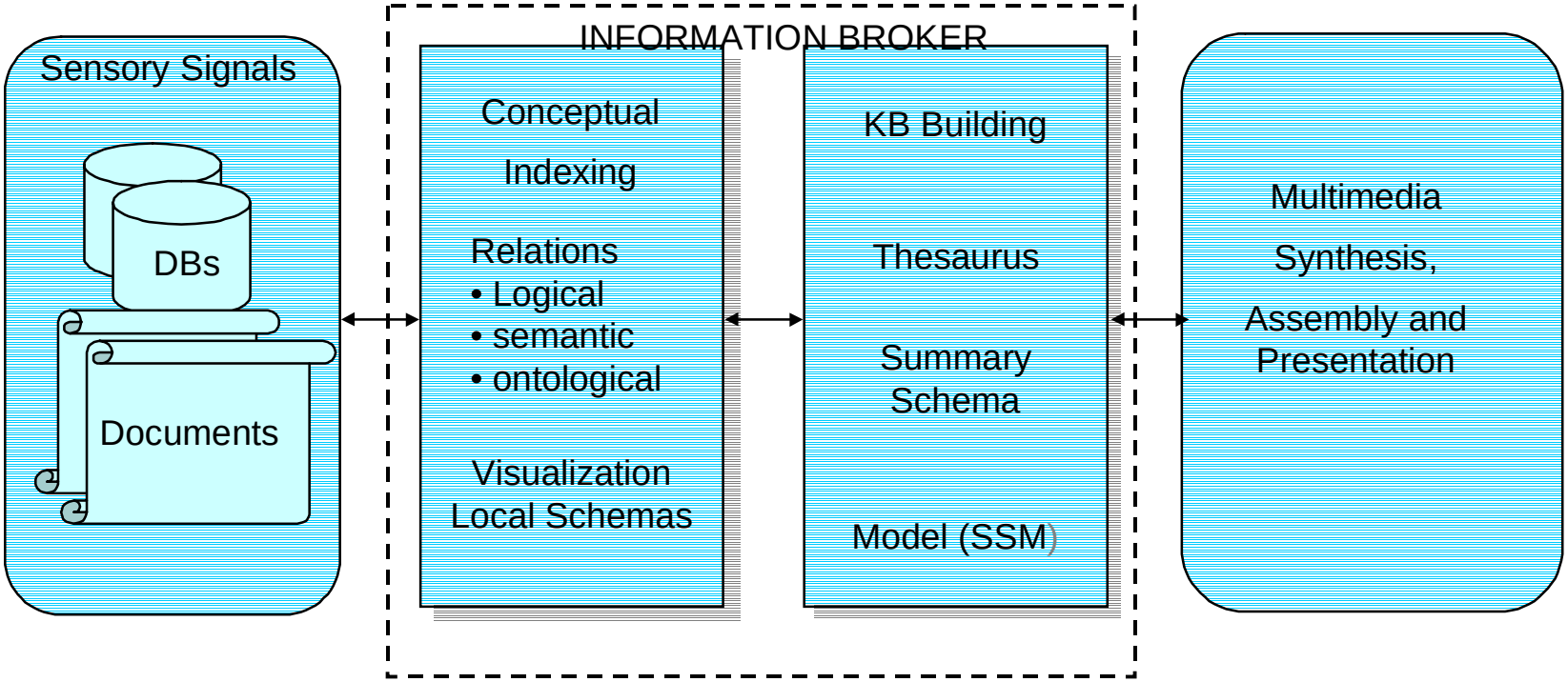
PENNSTATE INFORMATION BROKERING FOR KNOWLEDGE DISCOVERY

INFORMATION SOURCES
- DISPERSED
- TRANSIENT
- HETROGENOUS
- EMBEDDED IN AUTONOMOUS LEGACY SYSTEMS

KNOWLEDGE BASE DEVELOPMENT
- RULE BASED EXPERTISE
- SEMANTIC RELATIONSHIPS
- FUNCTIONAL RELATIONSHIPS
- SEMIOTIC LEARNING
- KNOWLEDGE ADAPTATION
- PRODUCT AND PROCESS MODELS
- OPERATIONAL GRAMMARS

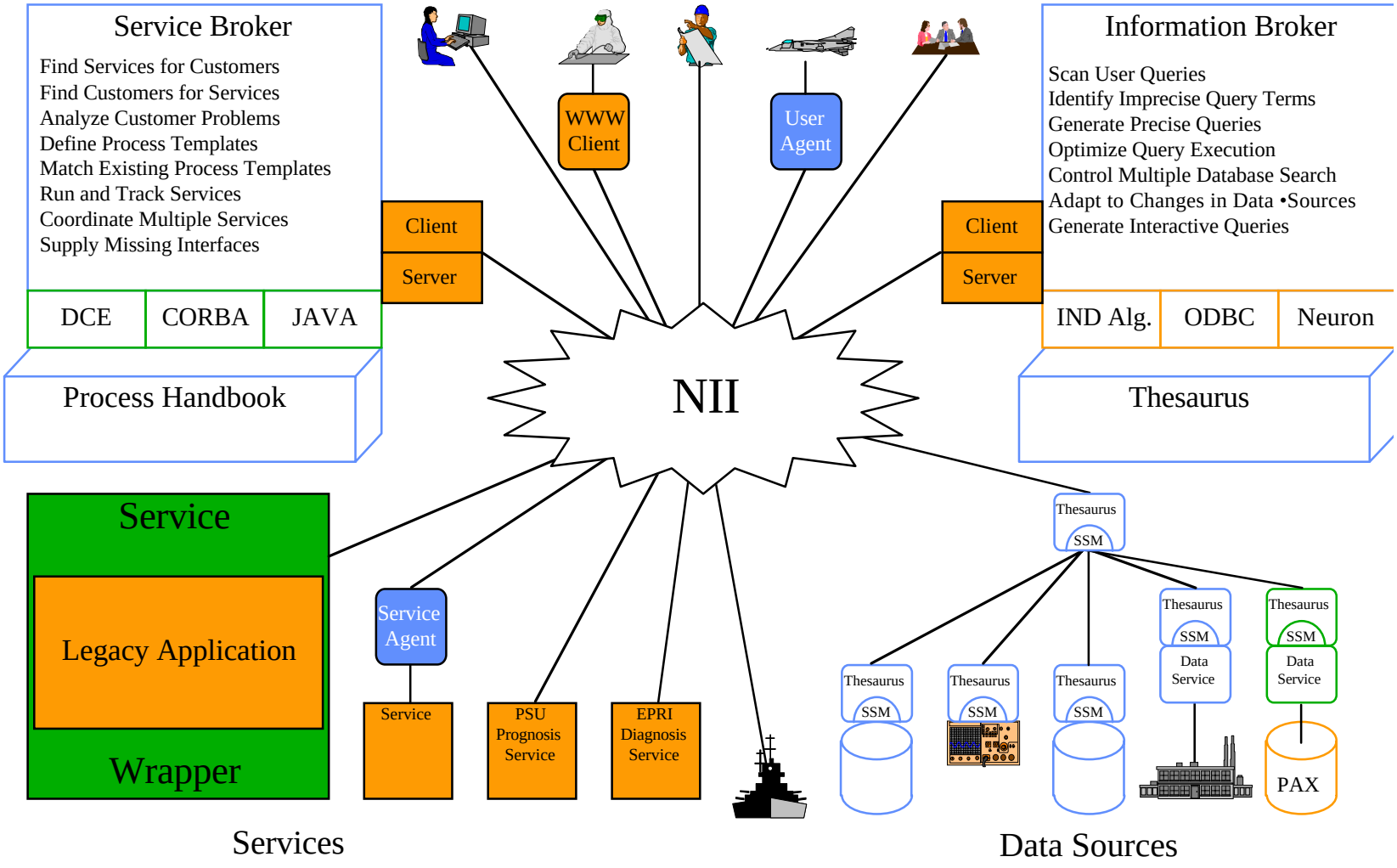
KNOWLEDGE DISCOVERY
- MULTI CONTEXT
- MULTIPLE AGGREGATION LVLS
- SEARCH MECHANISMS
  - PRECISE
  - BROWSE
  - FUZZY

USER NEEDS
- LOCAL AUTONOMY
- LOCAL SCHEMAS
- CONTEXT SPECIFIC SEMANTICS
- LOCAL CONSTRAINTS: TIME CRITICALITY, SYNCHRONIZATION
- INTEROPERABILITY WITH PRODUCT AND PROCESS MODELS, DEICISION TOOLS AND CONTROL MECHANISMS



PENNSTATE

DEFENSE MAINTENANCE INFORMATION INFRASTRUCUTRE



MATHEMATICAL MODELS OF SEMANTIC UNIFICATION

Semantic linking of personal knowledge allows interactive learning

What are these Semantic Structures?

- Semantic Networks [1]
- Associative Memories
- Additive Trees
- Hierarchical Clusters
- Pathfinder Nets
- Dimensional Structures

What are Knowledge Elicitation Methods?

- Semantic Similarity Ratings
- Free Word Associations
- Ordered Recall
- Free Recall
- Graph Building
- Indexing

Expert Knowledge vs.. Individual Knowledge Representation

- Mind Mapping
- Cognitive Learning

[1] Jonassen,D.H.,Beissner, K., and Yacci, M., “Structural Knowledge”, Hillsdale, N.J.,

INFORMATION ENTITY	DESCRIPTION	EXAMPLE
DOMAIN MODEL	(T,R) where T = set of all domain specific terms, R = set of all relationships between terms. Δ, R may represent a crisp or fuzzy relationship	Domain: Equipment Maintenance represented as a pathfinder semantic network [1]
QUERY	An expression in the domain model	SQL
SOURCE MODEL	Distributed models (T_e, R_e) representing contents of local information sources using a subset of domain terms and relationships	Set of local schemas with local node information domain specific terms [2]
SEARCH HIERARCHY	Semantic aggregation of terms in the source model using relationships in R and terms in T to create logical or physical nodes of a semantic tree network with a single root	Physical configuration (relationship: is a component of)
SEMANTIC DISTANCE	A function d: TxT Ψ [0,100] to represent semantic similarity of terms with regard to relations in R	Minkowski distance [1] Latent Semantic Index [3]
SEARCH ALGORITHM	Mechanism for searching the source model given a request for an expression in the domain model	A crisp network search algorithm [1]
ADAPTATION MECHANISM	Mechanism for updating the domain model, the source model and the associated semantic distances of terms	Initialize and update by usage

[1] Scvaneveldt, R.W., Pathfinder Associative Networks, Studies in Knowledge Organization, Norwood NJ, Ablex Publishing Corp, 1990

[2] Bright, M.W., Hurson, A.R., and Pakzad, S.H., "Automated Resolution of Semantic Hetrogeneity in Multidatabases", ACM Transactions in Database Systems, vol. 19, no. 2, pp. 212-253.

[3] Zha, H., "A Subspace-Based Model for Information Retrieval with Applications in Latent Semantic Indexing," CSE Technical Report 98-002, to appear in Lecture Notes in Computer Science, Springer-Verleg

METROLOGY FOR INFORMATION BROKERING

PERFORMANCE METROLOGY

Measures of Semiotic Intelligence

Semantic Intelligence

Precision

Recall

Rate of change of Precision Vs. Recall

Syntactic Intelligence

Semantic Inconsistency Measure

Contextual Independence Measure

Pragmatic Intelligence

Goodness of fit for Functional Relationships

Consistency of Semantic Aggregation

Structural Fitness Measures

Rate of change of Precision vs. Aggregation

ADAPTATION AND LEARNING MEASURES

Damping Sequences for Domain Adaptation

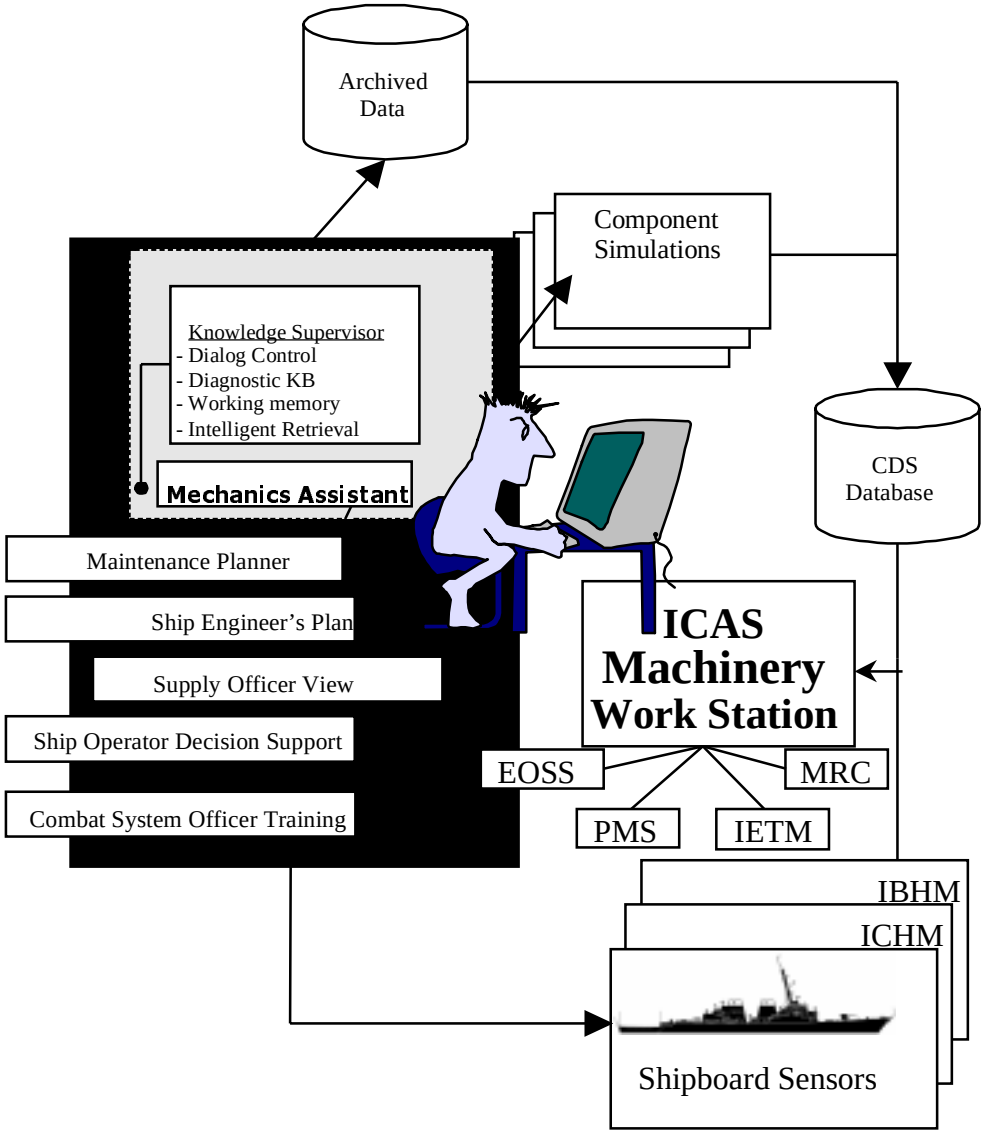
Usage Based learning

- Context adaptation through usage
- Learning new object relationships and adapting semantic similarity measures through usage

MEASURES OF COMPLEXITY OF THE KNOWLEDGE BASE

PENNSTATE

SHIPBOARD INTEGRATED LOGISTICS SYSTEM



EOSS - Engineering Operational Sequencing System

PMS - Preventive Maintenance Schedule

IETM - Interactive Electronic Technical Manual

MRC - Mobile Re-use Center

CDS - Configuration Data Sets

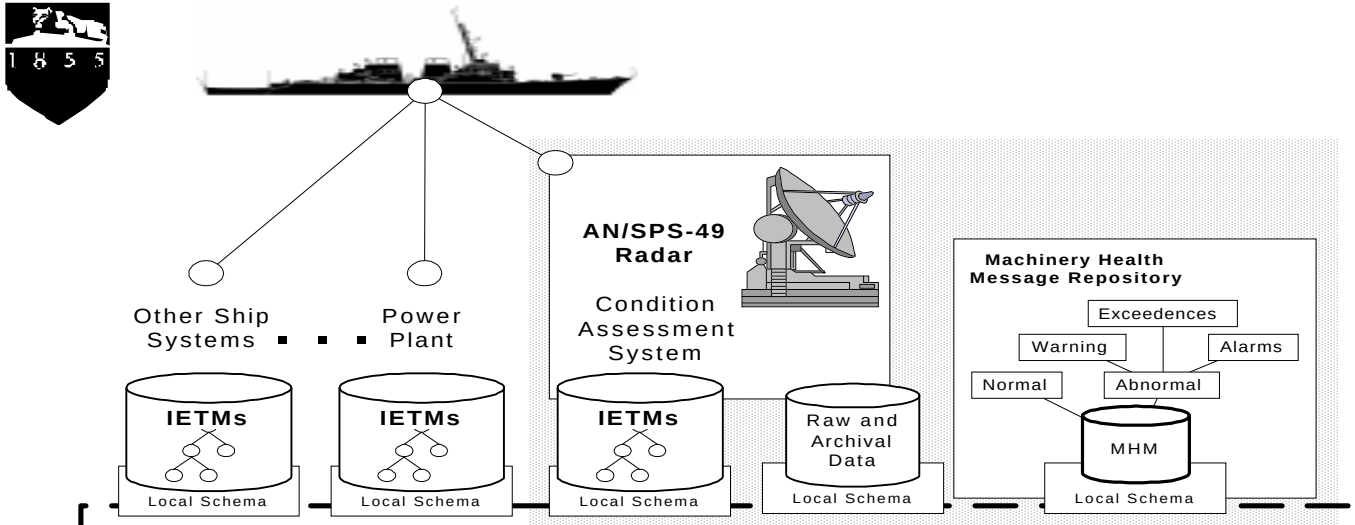
ICAS - Integrated Condition Assessment System

IBHM - Integrated Bearing Health Monitor

ICHM - Intelligent Component Health Monitor

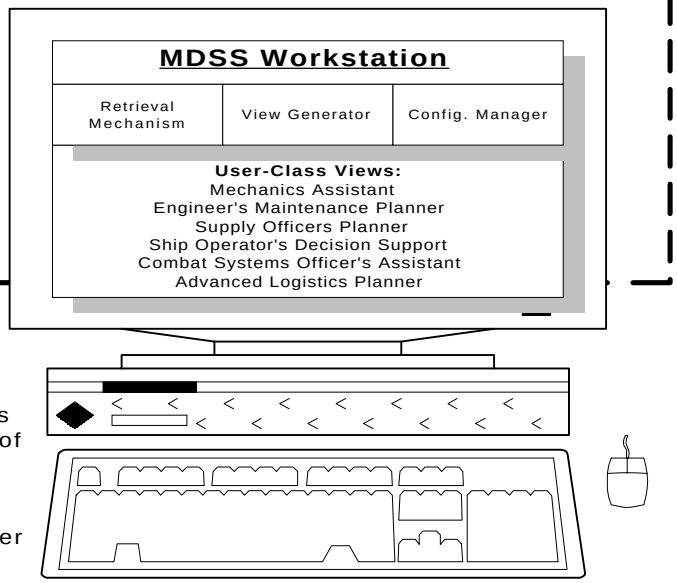
PENNSTATE

SHIPBOARD INFORMATION BASED MAINTENANCE



Intelligent Information Broker

- Integrate shipwide databases and services as nodes in maintenance-data source model
- Handle integration of legacy and new sources without changes to database
- Allow users to use local terms: intelligent interpretation of imprecise queries into domain terminology
- Incremental domain integration of ship components



- Intelligent MHM handling
- View specific decision support
- Intelligent connection between MHM and fault isolation procedures
- Access to information at any level of source model (any resource in the SSM tree)
- Interactive fault isolation
- Customizable tools tailorable to user needs

PRELIMINARY MEASURES OF SEMIOTIC INTELLIGENCE

SEMANTIC INTELLIGENCE MEASURES

Precision (Q) = 100* Φ (Y $\geq$ Z) / Φ (Y)

Recall (Q) = 100* Φ (Y $\geq$ Z)/ Φ (Z)

where Φ is a measure on information sets and for a query Q

Y(Q) = set of all relevant information in response to Q

Z(Q) = set of all retrieved information in response to Q

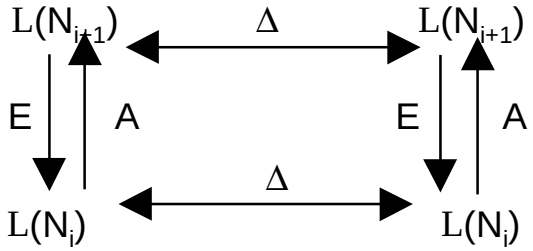
SEMANTIC UNIFICATION MEASURES

For a knowledge domain (T,R) organized as a hierarchical semantic network of levels i, combining terms at each lower level according to a formal grammar to generate languages ($L(N_i)$ of increasing abstraction and decreasing precision, define

A: $L(N_i)$ Ψ $L(N_{i+1})$ as the semantic aggregation function,

E: $L(N_{i+1})$ Ψ $L(N_i)$ as the semantic expansion function.

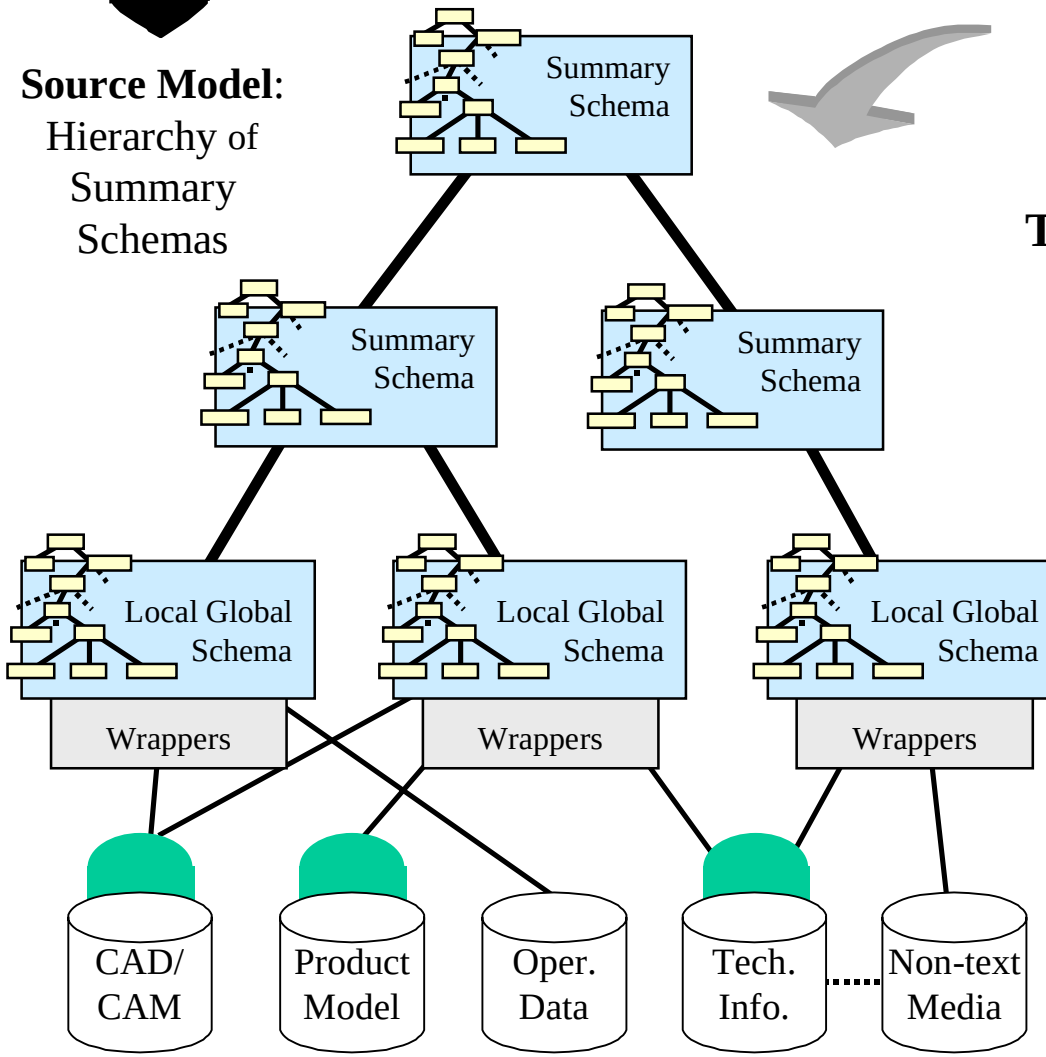
Δ , R is defined to be **semantically consistent** if the following diagram is commutative



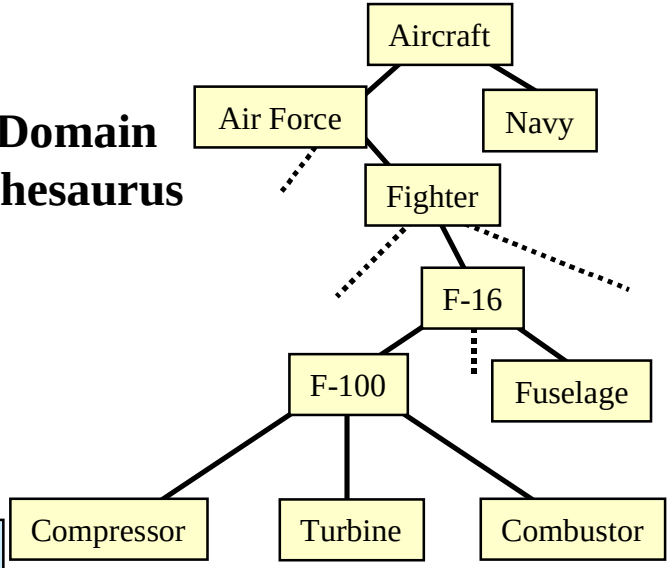
PENNSTATE

KNOWLEDGE BASE DEVELOPMENT

Source Model: Hierarchy of Summary Schemas



Domain Thesaurus



- **Electronic Knowledge representation**
  - **metaknowledge of dynamic adaptation to source data**
  - **tolerance of semantic heterogeneity of user communities**

PENNSTATE

SEMANTIC DISTANCE IN SHIP MAINTENANCE THESAURUS

Based on Pathfinder Networks

1) If Δ , R $\}$ t Δ t' then d $\}$ t Δ_d t'
(that is, a distance is assigned to every relational link in {T,R}

2) For each path = $\{t_1,\ldots,t_n\}$ in {T,R} from t_1 to t_n,
let $\{d_1,\ldots d_{n-1}\}$ be given by $t_i \Delta_{di} t_{i+1}$ for i = 1,…, n-1.

3) Define $d(t_1,t_2) = \min (\Sigma_{i=1,n} d_i^r)^{1/r}$
B , {T,R}

4) PFNets initialized by 1st response data from human subjects; mdp links initialized by default distances associated with each relation.

5) The distances d_i between related terms are subject to adaptation as users reinforce or suppress individual links or entire paths.

ADAPTATION OF SEMANTIC DISTANCE IN SHIP MAINTENANCE THESAURUS

Dampening Sequence - a sequence $\{\gamma_\vartheta\}$ of real numbers satisfying

$\lim_{\vartheta \to \infty} \gamma_\vartheta = 0$ (stability)

$\sum_{\vartheta=1}^{\infty} \gamma_\vartheta = \infty$ (independence from the past)

The adaptive distance d_i of a direct relational link (t,t')is

$d_i(1)$ = relation based default

$d_i(\vartheta) = (1-\gamma_\vartheta) d_i(\vartheta-1) + \gamma_\vartheta d^*$, where $d^* = d_{min}$ (reinforce)

d_{max} (suppress)

and

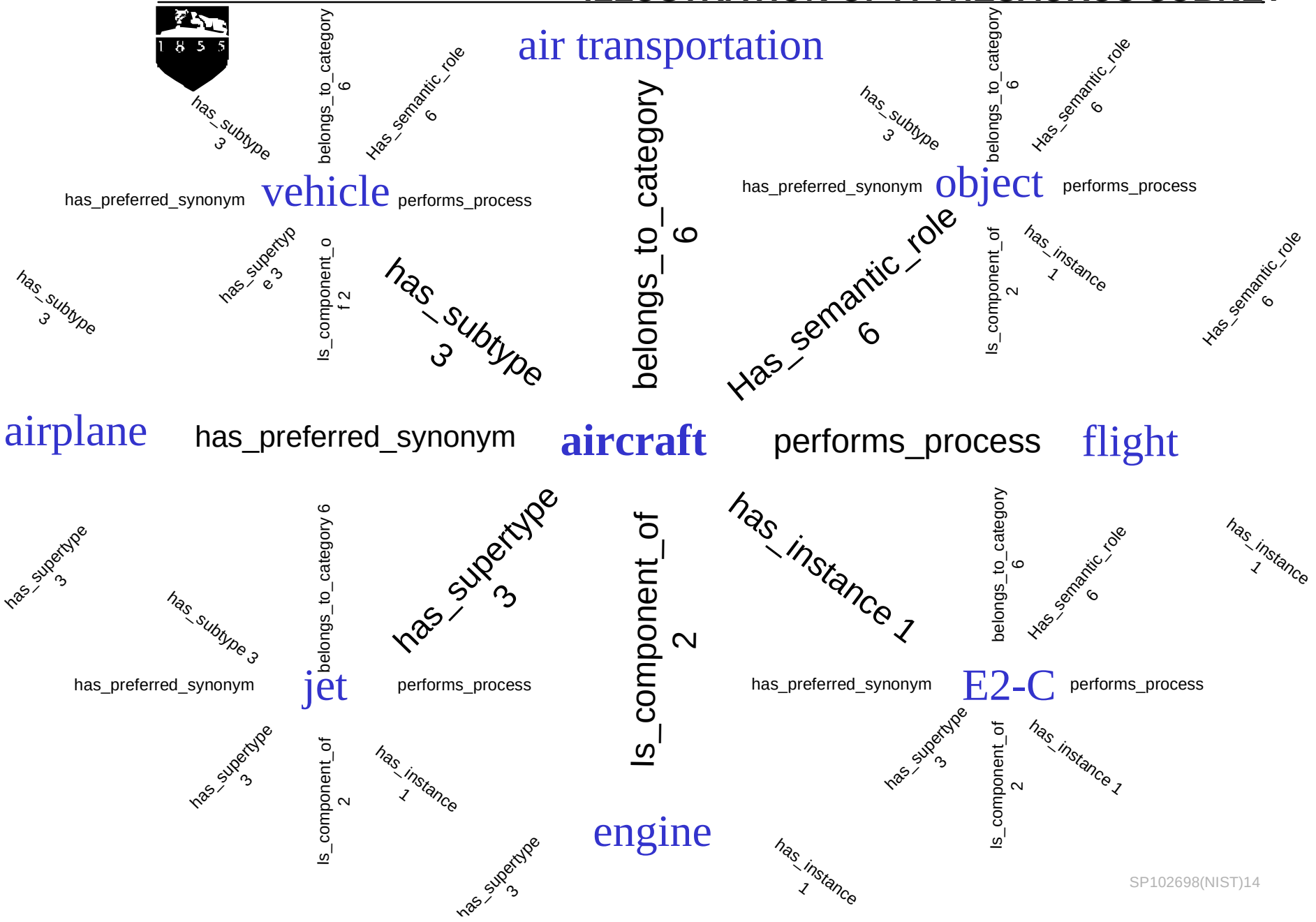
ϑ = number of times the particular link (t,t') has been updated by the particular user invoking the update.

Adapted link weights are stored in a profile for each individual user, and used only for sessions under his/her user_id.

$\gamma_\vartheta = 1/\vartheta$ is the current damping sequence in use; parameterized to facilitate future sensitivity studies.

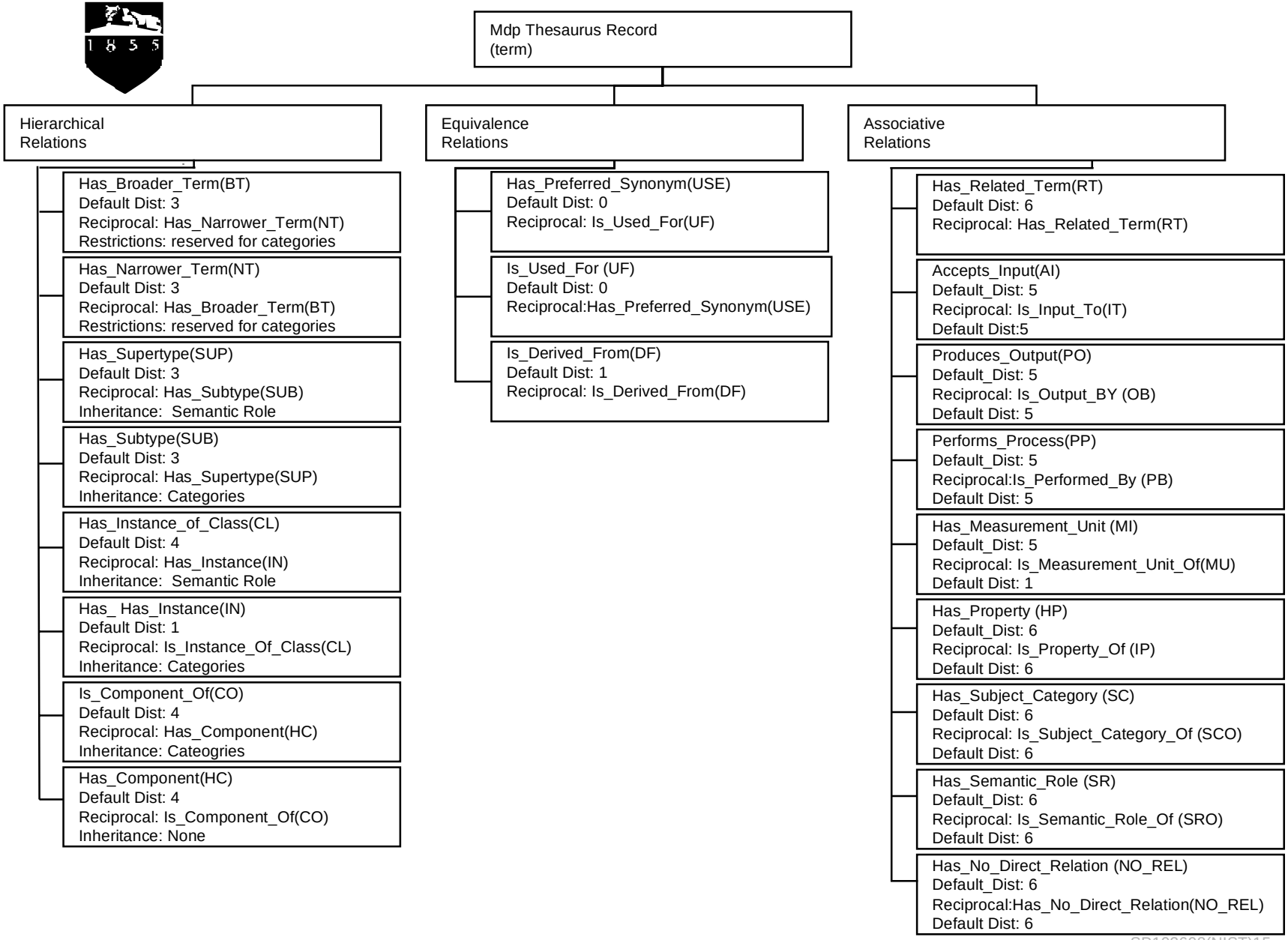
PENNSTATE

ILLUSTRATION OF A THESAURUS SUBNET



PENNSTATE

THESAURUS TERM ENTRY



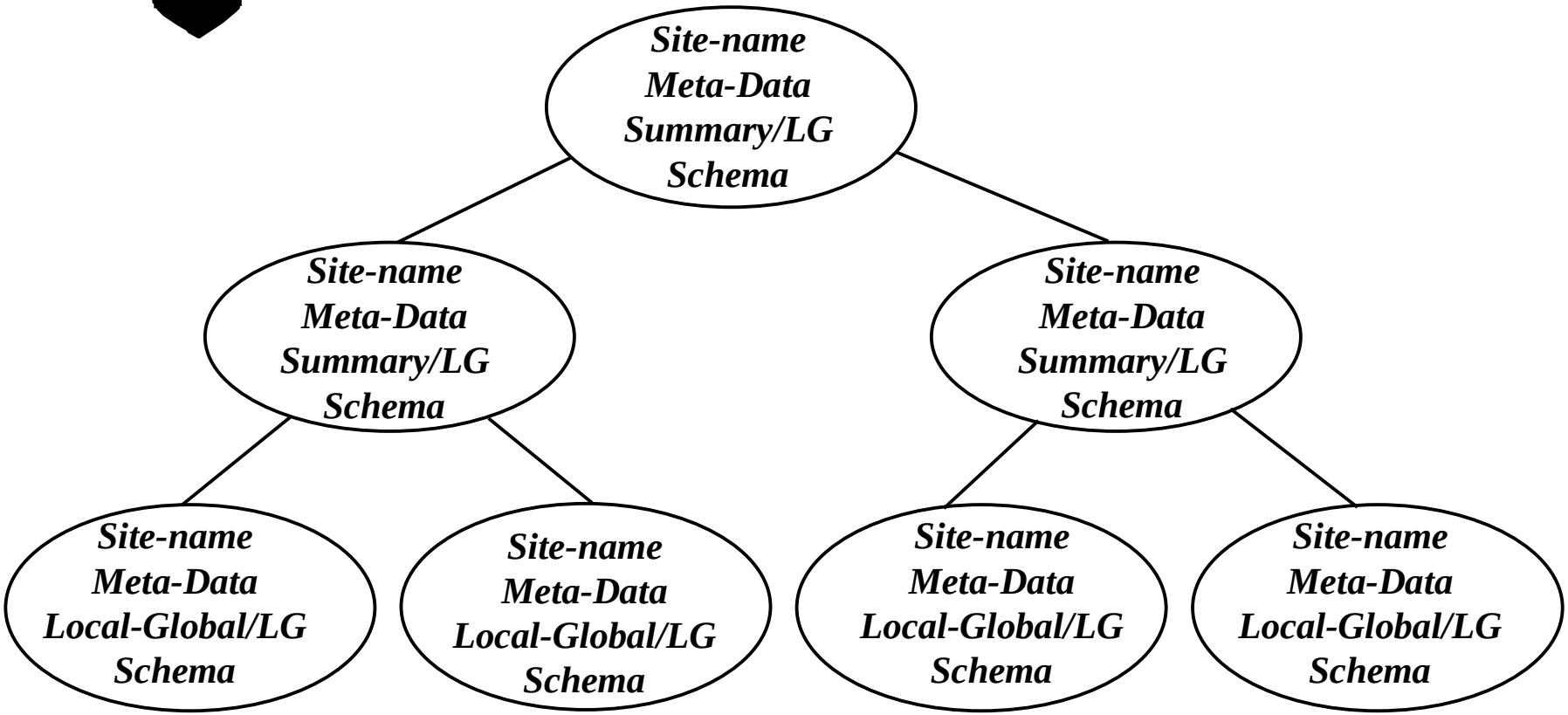
PENNSTATE

SOURCE INFORMATION MODEL IMPLEMENTATION (DESIGN GOALS)

- hierarchical semantic network
- preserves local autonomy of information sources
- preserves user perspective
- supports efficient search and retrieval
- dynamic organization; flexible and scalable

PENNSTATE

SUMMARY SCHEMAS NETWORK STRUCTURE



Database

Electronic manual

Simulation/ automated analysis

CORBA objects

PENNSTATE

SUMMARY SCHEMAS NETWORK DYNAMICS

Prerequisites - node daemons running on all host computers
- local-global schema/meta-data files on all leaf node hosts
- servers running (thesaurus, network admin, query generator)

Pre-Configure - User Network Admin GUI to build from scratch or modify a saved graph and set node parameters such a log file activation.

Initiate Build - Admin GUI sends
start_node messages to node daemons
parameter messages to appropriate nodes
new-parent messages to leaf nodes

Self Propogating Configuration -
nodes send messages to parents announcing themselves and forwarding their schemas

parent nodes request common ancestors of terms and categories from all child schemas and construct summary schemas

upon every change, a node sends its new schema to its parent

upon receiving a new parent message, a node notifies the old and new parents of the change

Message Handler

Site-name:	AN/PSQ-6 Radio IETM
Parent Node:	Navy
Categories:	Navy Watercraft
Access:	**IETM Interface Specification**
site-info:	"This site contains IETM #i2778517348 for the AN/PSQ-6 Radio. The manual is organized by radio components."
LG-Schema:	
Radio	Radio_Set
Receiver	Receiver Transmitter
Transmitter	Receiver Transmitter
Interface	Interface Module
Hybrid Unit	Basic Hybrid Unit
Controls	Control Panel
Power	Power Switch
Salinity Cell	Interface Module Salinity Cell

PENNSTATE

ANATOMY OF A PARENT NODE SUMMARY SCHEMAS NETWORK

Message Handler

Node Name: Navy

Parent Node: Global

Categories: Navy Aircraft, Navy Watercraft

Summary Schema:

aircraft engine:	child node i_1, child node i_2, …,child node i_k
radio	child node j_1,child node j_2,…, child node j_m

PENNSTATE

ANATOMY OF AN IMPRECISE QUERY

search terms - one or more thesaurus terms

semantic distance - tolerance for ‘matching’ related terms

category - restricts search to one or more thesaurus categories

site - site (node) name defines the subtree at which search begins

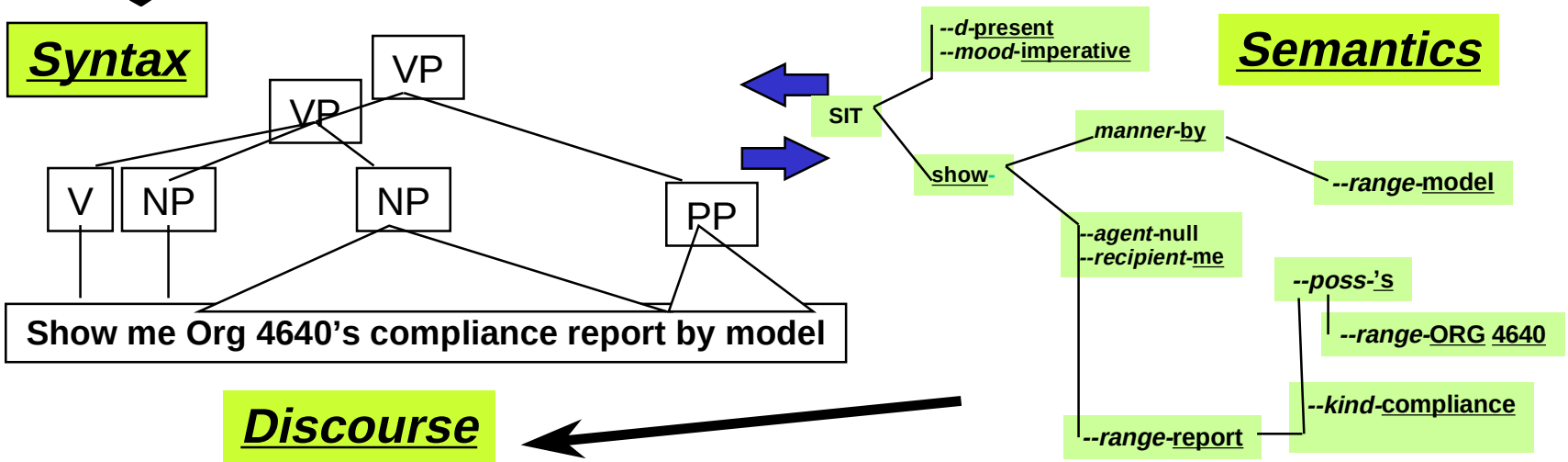
id - for tracking, assigned automatically by the query server

user_id - for adaptative thesaurus

PENNSTATE

ANATOMY OF A QUERY RESOLUTION

INPUT QUEIRES UNDERGO SYNTACTIC, SEMANTIC & DISCOURSE



Discourse

```
query
        name:               q_1
        msg_id              'sent-8'
        q_type:             retrieve
        object:             e_1
ent
        name:               e_1
        msg_id:             'sent-8'
        sem_category:       report
        report_type:        e_2
        major_grouping:     e_3
ent
        name:               e_2
        msg_id              'sent-8'
        sem_category:       'schedule compliance activity'
ent
```

Information Broker

PRECISION VS. RECALL TRADEOFFS

"A General Framework for Computational Intelligence", Paul S. Prueitt, NIST ISAS 98, 15 Sept 1998

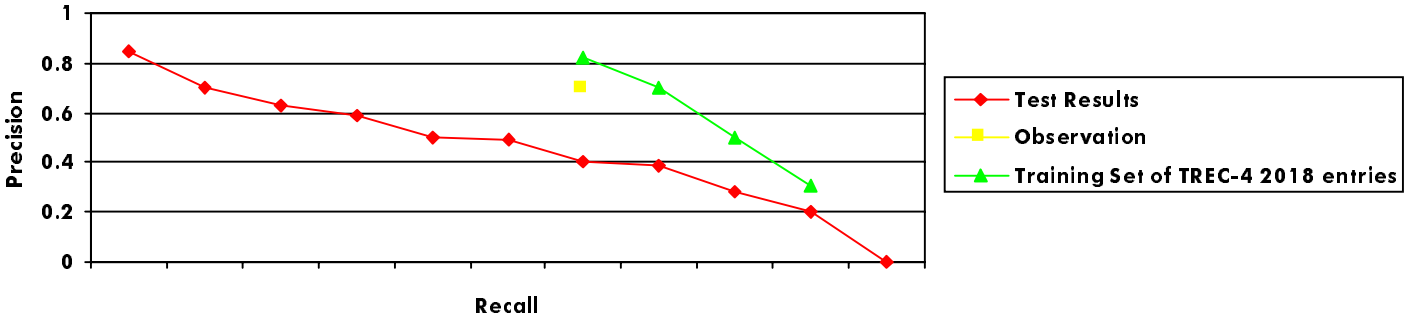
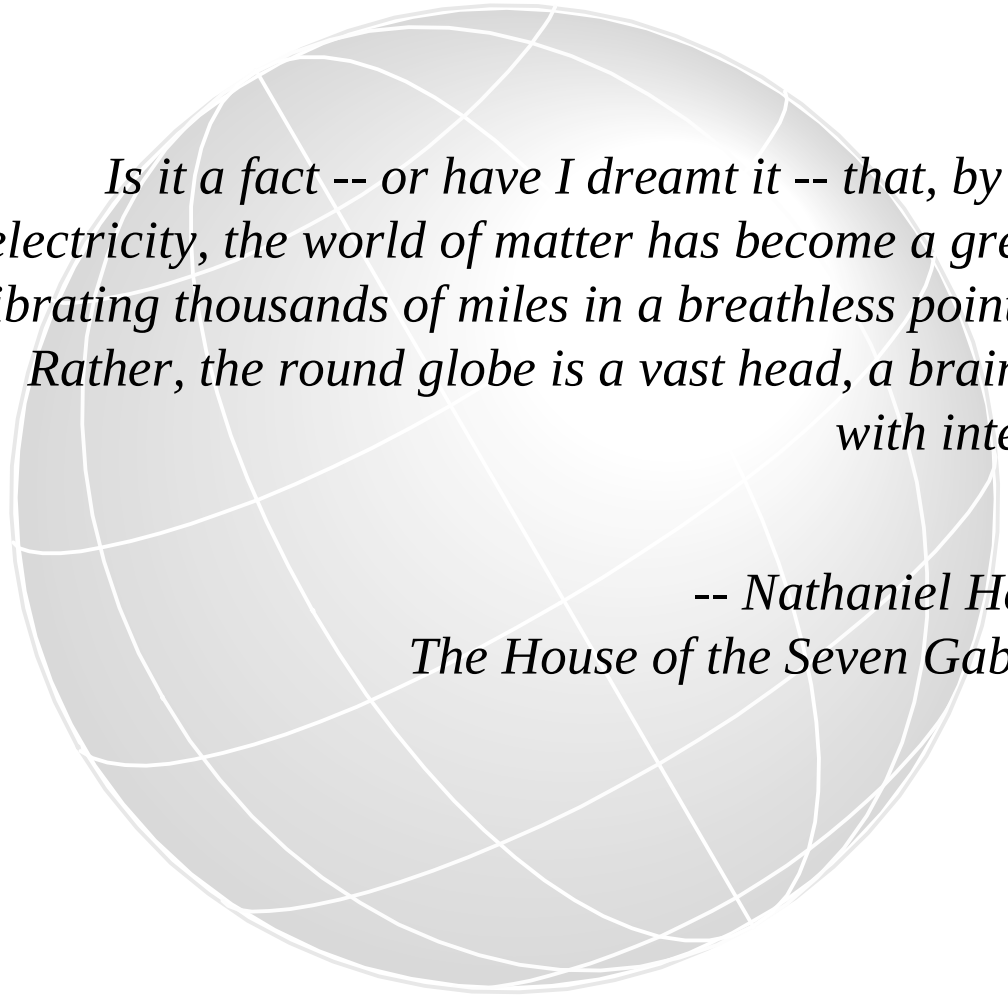


Figure 1. TREC-4 Wall Street Journal Collection 4556 entries with topic assignments 174,000 full set

OPEN ISSUES

- Computational grammars for adaptation
- Semantically consistent adaptive memories
- Limits of adaptation for semantic fitness
- Exception Handling
- Inferencing from imprecise semantic expressions
- Knowledge discovery methods
- Standardized test data sets

MEASURES OF COMPUTATIONAL INTELLIGENCE



Is it a fact -- or have I dreamt it -- that, by means of electricity, the world of matter has become a great nerve, vibrating thousands of miles in a breathless point in time? Rather, the round globe is a vast head, a brain, instinct with intelligence!

-- Nathaniel Hawthorne
The House of the Seven Gables, 1851